Host-directed vibroacoustic biosignature of viral respiratory infection

Andreas Schuh, *Member*, *IEEE*, Michael Morimoto, Piotr Kaszuba, Kevin Hammond, Jerry Swan, Krzysztof Krawiec, NL Shasha Jumbe, *Member, IEEE*

***Abstract*— Respiratory infection testing, patient risk, and clinical costs are driven by disease prevalence, direct and downstream costs of false-positive/negative tests. In addition, false-positives decrease with disease prevalence, while -negatives increase. We developed a host response-based, easy to use, fast, low-cost, non-invasive respiratory infection detection system and used COVID-19 as a mechanism for assessing host vibroacoustic biosignatures based accuracy, sensitivity, and specificity to distinguish SARS-CoV-2 carriers from non-carriers in a diverse US, Indian, and African population.**

***Clinical Relevance*— A vibroacoustic infrasound-to-ultrasound e-stethoscope that can detect a respiratory infection early in both asymptomatic and vaguely symptomatic, like cough, patients that could be a tool to meet current as well as future needs as viral pathogens come and go, seasonally and within pandemics.**

I. Introduction

Screening tests are not considered diagnostic, but are primary tests used to flag individuals with high probability of disease that should undergo additional testing in order to accurately establish the presence or absence of disease [1,2]. Acoustics have long been at the center of patient care, with the stethoscope emerging as a timeless symbol of the caregiver–patient relationship. In many ways, the stethoscope has remained largely unchanged since Laënnec [3,4]. However, greater than 90% of low frequency and low amplitude data emitted by the throat, heart, lung, and digestive tract, is below the human threshold of hearing. These data are ignored as "noise", are unobserved, and/or are uncollectable by current stethoscopes [5-8].

Our overarching hypothesis is that known audible sounds combined with previously unobserved low frequency and low amplitude vibroacoustic data, inaudible to the human ear and below current instrument detection, could discriminate between confirmed SARS-CoV-2 carriers and non-carriers. The proof-of-concept pilot study hypothesis included audible (20 Hz – 20,000 Hz) sound combined with inaudible (< 20 Hz) infrasound that could discriminate between confirmed COVID-19 patients and matched patients who do not have lung-related diseases. The subsequent multinational study hypothesis was that the COVID-19 vibroacoustic biosignature extracted from the matched case-control pilot study could be generalized to participants with a variety of ethnic backgrounds, body types, and ages suspected of SARS-CoV-2 infection.

In this report, we introduce a novel host response-driven COVID-19 vibroacoustic biosignature Software as a Medical Device screening system. The hardware comprises of an easy to use, fast, non-invasive, broad-frequency spectrum e-stethoscope. Collected vibroacoustic data, clinical symptom data, contextual (device orientation and environmental determinants) data are de-identified, labelled, encrypted, compressed, and subsequently analyzed using transparent, interpretable, and auditable algorithms for disease classification. The software can be updated to incorporate real time disease prevalence to adaptively minimize test false-positives/negatives. Our main finding was the discovery that a host response-based COVID-19 vibroacoustic biosignature can be used for accurate detection of respiratory infection early in both groups who are asymptomatic or who report a vague symptom, like cough.

II. Methods

A. Infrasound-to-Ultrasound voice coil transducer e-stethoscope

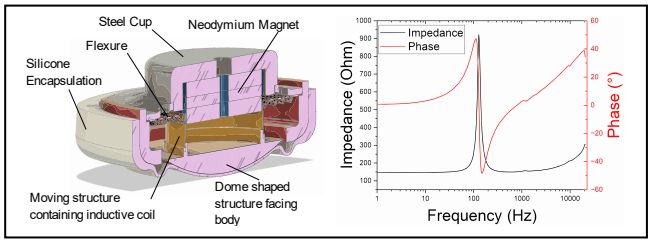

Figure 1. imPulse™ UNA vibroacoustic chestpiece (left) and response (right).

The imPulse™ UNA e-stethoscope is based on a voice-coil transducer, also known as a moving-coil dynamic transducer (Figure 1). The imPulse™ UNA infrasound-to-ultrasound vibroacoustic e-stethoscope chestpiece collects audible sounds and inaudible vibrations generated by the body. It includes a transducer subsystem capable of detecting low amplitude vibration and acoustic signals below the human-audible threshold infrasound (0.1 Hz – 20 Hz), human audible-frequencies (20 Hz – 20 kHz), and higher-inaudible frequencies (20 kHz – 24 kHz). Voice-coil transducer actuator systems are currently used for a wide range of applications, including in optics and audio (microphones and speakers). The

*Research supported by Schmidt Futures and Uniformed Services University of the Health Sciences USUHS-4DBio³ Center for Biotechnology Prime Award Number: HU00012120093, Research Title: 4D Bio³ Fabrication in Austere Military Environments (FAME): Passive Detection SARs-CoV-2 Breathalyzer for Operational Medicine.

A. Schuh, M. Morimoto, NL Shasha Jumbe[#] are with Level 42 AI, Inc., Mountain View, CA 94041, USA; P. Kaszuba, K. Hammond are with Hylomorph Solutions Ltd., United Kingdom; J. Swan is with the University of York, United Kingdom; and K. Krawiec is with the Poznan University of Technology, Poland, and Hylomorph Solutions Ltd., United Kingdom.
[#]Corresponding author: +1 650-515-6112, e-mail: shasha@level42.ai.

chestpiece diaphragm is coupled with the patient's skin or placed over light clothing. As the diaphragm moves, so does the voice-coil. Because the chestpiece diaphragm moves back and forth, the magnetically induced current across the conductor and the voice-coil rises, falls, and changes polarity. This creates an alternating voltage across the coil, a direct electrical representation of the audible sound wave or inaudible vibration.

B. Clinical Study Design and Oversight

We trained, validated, and externally tested a machine learning system using demographic, clinical, and broad-spectrum infrasound-to-ultrasound (0.1 Hz to 24,000 Hz) vibroacoustic data trained against Reverse Transcription Polymerase Chain Reaction (RT-PCR) confirmed COVID-19 health status (Figure 2).

C. Study Population

A proof-of-concept, matched case-control, non-interventional, observational cohort study was performed in an over-characterized, hospitalized cohort to extract a putative SARS-CoV-2 infection vibroacoustic biosignature. The one-to-one matched case-control clinical design required one hospitalized control without COVID-19 diagnosis (RT-PCR negative) and no pulmonary diagnosis or symptoms to be matched to a confirmed RT-PCR positive symptomatic COVID-19 case. Matching criteria included age (± 5 years), gender, body mass index (± 15% BMI units), and history of smoking or vaping within the last 3 months (yes/no). The study consisted of two visits, separated by 2 (+/-1) days to assess stability of the vibroacoustic signal and any longitudinal changes associated with severity of the disease.

The subsequent multinational, multi-center, multi-ethnic study recruited (a)symptomatic walk-ins and outpatient participants from representative populations in India and République du Congo who were being screened for COVID-19 infection during and across wild type, delta, and omicron variant outbreaks. RT-PCR negative controls were at risk of becoming cases and came from populations with the same exposure distribution as the cases.

D. Vibroacoustic Data Acquisition

In the proof-of-concept study, subjects were required to undergo an anterior-only auscultation exam overlapping with Volpicelli [9] lung zones across six body posture/positions (neutral sit, neutral upright stand, supine, left lateral decubitus position, handgrip and cough, and vocalizations that reveal pulmonary consolidation), see Figure 2. In each posture position, the chestpiece was held for 30 seconds at each auscultation point following a 9-step staircase ladder procedure, starting at the right carotid. Based on insight gained from the pilot study, data collection for the multinational study was simplified by focusing on neutral sit and neutral upright stand postures/positions only in order to reduce exam times for scalability. Off-body cough data collection was added to the multinational study to replicate voice-as-biomarker smartphone app data acquisition [10-14].

E. Outcomes: Definition of COVID-19 infection status

Cases were defined as participants with a positive COVID-19 RT-PCR test determined to achieve 95% or greater positive percent agreement (PPA) between test results and attest results from an FDA Emergency Use Authorization comparator (>95% PPA EUA RT-PCR) as the gold standard detection method for the diagnosis of SARS-CoV-2 infection.

III. DEVELOPMENT OF THE CLASSIFICATION MODEL

A. Structural Machine Learning (SML) classification method

Each participant's data was comprised of demographic information, clinical symptoms, vibroacoustic data across body posture, auscultation site, and RT-PCR COVID-19 status. The vibroacoustic recordings were down-sampled and trimmed into 10-second non-overlapping segments, which were further split by frequency band. These vibroacoustic segments were used as input into machine learning models implemented as Domain-Specific Language (DSL) computer programs. DSLs are a means of representing expertise in machine-executable form. Signal processing and feature extractor DSLs were used to make RT-PCR COVID-19 status predictions to generate COVID-19 vibroacoustic biosignatures. To combat training-set overfitting, SML classification was based on aggregation of multi-level classifier decisions that used majority voting. Level 1 classifiers (including logistic regression, multilayer perceptron, random forest, linear SVM, radial SVM, etc.) used individual and ensemble classifiers to map identified features and interactions between features. Each Level 2 classifier voted on top of an ensemble of Level 1 classifiers. The ultimate Level 2 classifier DSL queried Level 1 classifiers, in order to aggregate decisions and produce the final decision report (probabilities, feature values, and execution trace).

B. Deep Learning classification method

State-of-the art deep neural networks/deep learning (DL) models were applied, comprising of three DL architectures designed to learn from time-series data: 1) Long Short-Term Memory models (LSTM), 2) Convolutional LSTMs (ConvLSTM), and 3) models based on transfer learning from pretrained audio classification models (e.g., YAMNet).

C. Comparison of Models

An illustrative, rather than definitive, experiment compared DL models against SML in a blinded-out-of-sample evaluation.

D. Measuring Model Performance

We used Shannon entropy to determine the balance of COVID-19 positive to negative cases in each sub-study. AUC was used as an overall measure of class discrimination. In addition, sensitivity, specificity, and accuracy were calculated. Bootstrapping was used to estimate 95% confidence intervals (CIs). AUCs were compared using the nonparametric Delong method [15].

E. SML Model Optimization

To ensure algorithm robustness, we assessed SML system performance 1) by extracting and quantifying the importance of auscultation site and body posture/positions, 2) with and without biased self-reported COVID-19 symptoms, and 3) by optimizing the class discrimination threshold (decision threshold) by which SML continuous probabilities output were converted into COVID-19 positive/negative class labels.

Threshold optimization was performed on the out-of-sample test set of the multinational study model by scaling the confusion matrix entries to reflect simulated disease prevalence. The best thresholds for each disease prevalence scenario were used to calculate 95% confidence intervals using bootstrap resampling.

F. Model overfit mitigation

To minimize data leakage and overfitting, training and testing datasets were created by stratified (class balance-preserving) sampling. In the pilot study, test datasets remained at rest behind a firewall and were accessed blindly to a) make data leaks impossible, b) reduce overfitting risk (where models fail to generalize to the test data), and c) closely reflect real-world deployment. For COVID-19 vibroacoustic biosignature testing in the multinational study, out-of-sample test datasets were utilized in periodic evaluation with the mindset to minimize data leaks and overfitting.

IV. RESULTS

A. Characteristics of the Data Sets and Studies

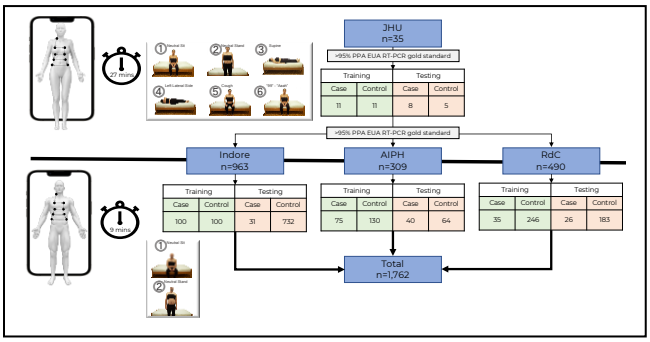

Figure 2. Enrollment of case and control subjects for the training and test data sets in the matched case-control proof-of-concept pilot study to extract the putative vibroacoustic biosignature and the multinational, multisite, multiethnic clinical study. JHU=Johns Hopkins Hospital and Johns Hopkins Bayview Medical Center; Indore= Central Labs, Indore, India; AIPH=Asian Institute of Public Health, Bhubaneswar, India; RdC=Labo HDL (Fondation Marie Madeleine Gombes), Pointe-Noire, République du Congo.

The pilot case-control study included a total of 35 inpatients [20 female and 15 male; 32.4 (19.2 – 44.3) kg/m2 BMI; 48.6 (30-78) years of age], who each provided 27-minute exam deep-phenotype vibroacoustic data. 19 hospitalized adult men and women over 18 years of age with a positive RT-PCR test for SARS-CoV-2, collected from a respiratory sample within the last 7 days and pulmonary symptoms within 72 hours of enrollment not receiving ventilator support (cases), were first enrolled. Cases were over-enrolled because of challenges finding matched controls. 16 case-matched inpatients with a negative RT-PCR test for SARS-CoV-2 and no pulmonary diagnosis or symptoms (controls) were subsequently enrolled. By design, the matched case-control study was well balanced (Shannon's entropy = 0.994). In the hospitalized inpatient pilot study, the threshold-optimized SML system discriminated COVID-19 negative controls from positive cases with an AUC of 1.0, a sensitivity of 100% (95% confidence interval not possible), and a specificity of 91.6%.

B. SML-vs.-DL Comparison of models

We compared our SML system to a range of state-of-the art DL approaches, including convolutional Long Short-Term Memory (LSTM) [15] and pretrained YAMNet-based models (MobileNetV1 provided in TensorFlow) [16]. Both the DL and SML models achieved a perfect ROC AUC score of 1.0 on the training dataset. When assessed on the blinded out-of-sample test set, the best DL model achieved a ROC AUC of 0.752 and the unoptimized SML system achieved 0.837. This result suggests that SML is less prone to overfitting than DL models for small training cohorts.

C. Multinational Out-of-sample Test Data Set

A total of 1,762 multinational study participants (1,455 COVID-19 negative controls, 307 COVID-19 positive cases by RT-PCR; 642 female and 1,120 male; 24.1 (21.7 – 26.9) kg/m2 BMI; 36 (27-50) years of age), provided 9-minute exam vibroacoustic data per subject. The pilot SML system was first re-trained and validated on 686 participants (476 COVID-19 negative controls, 210 COVID-19 positive cases by RT-PCR collected across 695 auscultation sessions) to transition from the 27-minute pilot study "deep-phenotype" to the 9-minute "shallow-phenotype" vibroacoustic biosignature. A separate out-of-sample test dataset of 1,076 multi-country, multi-center study participants, including 979 COVID-19 negative controls and 97 COVID-19 positive cases by RT-PCR, was used for model testing. As expected, the multinational, multi-center sub-studies were not well balanced. Shannon's entropy ranged from 0.540 to 0.953, where COVID-19 positive cases ranged across sub-studies from 12.5% to 50%. The SML system had an out-of-sample test set AUC for the detection of COVID-19 of 0.88 (95% CI, 0.84 to 0.91), a sensitivity of 76.5% (95% CI, 67.3 to 84.7), and a specificity of 82.8% (95% CI, 80.5 to 85.1).

D. Software as a Medical Device Model Calibration

A cumulative learning curve was generated to help determine model underfit or overfit to demonstrate the adequacy of training and testing datasets for the classification task. Model transparency, interpretability, and an algorithm audit trail was examined by tracing vibroacoustic biosignature features across matched case-control pairs across auscultation site and body posture.

Since self-reported symptoms may be under- or over-reported, we distorted patient reported symptoms by shuffling (randomly swapping symptoms between sessions for a given patient) or biasing (permanently switching the self-reported symptom to the opposite for 10, 20, and 50% of the patients). Distorted average out-of-sample test set AUCs fell below 50% following self-reported symptom input, while the SML alone and SML-plus-symptoms system was resilient to adversarial input.

Lastly, the SML system was configured to be a real-time adaptable solution that reduced the probability of false-positive and false-negative results and downstream public health costs with varying disease prevalence. Optimized positive/negative class decision thresholds were determined which maintained AUC for the detection of COVID-19 of 0.88 (95% CI, 0.84 to 0.91) and vibroacoustic biosignature test accuracy ranging from 99.5% (95% CI, 99.5 to 99.5) to 81.5% (95% CI, 78.8 to 84.1), at 0.5% through to 25% COVID-19 disease prevalence, respectively.

V. DISCUSSION

The strengths of the reported vibroacoustic biosignature study are four fold: a) focus on a host response-based screening test avoids variant-of-concern changes that potentially impact molecular, antigen, and serology test performance; b) the screening system was trained and tested to discriminate COVID-19 positive cases versus COVID-19 controls as confirmed by RT-PCR gold standard molecular tests; c) design of experiments, in which a generalizable COVID-19 screening test signature was extracted in a well-characterized cohort for subsequent validation in a heterogenous multiethnic population; and d) the study deliverable outcome was a tunable system that maximized test accuracy according to disease prevalence for more efficient allocation of limited public health resources to improve the quality of care for patients. Our main discovery was a COVID-19 vibroacoustic biosignature that can be used for easy, fast, non-invasive, and accurate detection of symptomatic and asymptomatic SARS-CoV-2 carriers based on novel vibracoustic data.

Our study has limitations. More training data creates more performant algorithms and this effort is no different. Notably, our study was prospective. Data was collected prospectively over a period of several months from a number of multinational centers during and across wild type, delta, and omicron variant outbreaks. Host and virus genome-wide association studies to characterize human genetic determinants and variant differences on impact of COVID-19 susceptibility and severity in the multinational study are ongoing in a limited number of collected samples.

Several studies have suggested that COVID-19 screening and detection can be performed by smartphone-based voice and cough applications [10-14]. Notably, smartphones are optimized to cover audio frequencies ranging between 50 Hz to 7,000 Hz [17]. In contrast to our study, these reports are largely based on patient-reported data, lack characterization and standardization of data collection equipment, and did not use RT-PCR as the clinical gold standard. The proposed vibroacoustic biosignature combines broad-bandwidth audible and inaudible acoustic data and efficient algorithms as a potential new tool in the healthcare worker's toolkit for the advancement of the continuum of screening, disease detection, diagnostics, and monitoring on therapy.

VI. CONCLUSION

We developed a host response-based, easy to use, fast, low-cost, non-invasive respiratory infection detection system and used COVID-19 as a mechanism for assessing host vibroacoustic biosignatures accuracy, sensitivity, and specificity. The resulting screening system was trained and tested on a large and ethnically diverse population and had high sensitivity and specificity for discriminating between COVID-19 positive cases versus COVID-19 controls as confirmed by RT-PCR. Further investigation is in progress to prospectively validate the tunable system across COVID-19, respiratory syncytial virus, and influenza A/B [NCT05765396], and tuberculosis [NCT04923958].

ACKNOWLEDGMENT

The authors want to first acknowledge the many lives lost or changed by the COVID-19 pandemic. Second, we would like to acknowledge and thank all study participants, clinicians, and staff in the US, India, and République du Congo studies for their trust, public health interest, and selflessness. Third, we are indebted to Kelly E. Dooley, Piotr Kaszuba, Hugues Loemba, Arjit Mohapatra, Margaret Krazne, Francis Olivier Tchikaya, Anicet Boumba, Edward Fuchs, Siddhanta K. Kar, Sarvjeet Nayak, Manas Swain, Ron Deiotte, Ryan Hafen, Bartosz Mikulski, Andrew Turner, Peter Rexelius, and Pinaki Panigrahi.

REFERENCES

[1] David L Sackett. Screening for early detection of disease: to what purpose? Bulletin of the New York Academy of Medicine. 1975; 51(1):39.

[2] Cheryl Herman. What makes a screening exam" good"? AMA Journal of Ethics. 2006. ; 8(1):34–37.

[3] I.R. Hanna and M.E. Silverman. A history of cardiac auscultation and some of its contributors. Am J Cardiol. 2002; 90-3, 259-267.

[4] David L, Dumitrascu DL. The bicentennial of the stethoscope: a reappraisal. Clujul Med. 2017;90(3):361-363.

[5] Leatham A. Auscultation of the Heart and Phonocardiography, 2nd ed, J and A Churchill, London 1975.

[6] Randall JM, Matthews RT, Stiles MA. Resonant frequencies of standing humans. Ergonomics. 1997;40(9):879-886.

[7] Eugenijus Kaniusas. Transmission of body sounds: an overview. Ultrasound. December 2005, 1(58):7-12.

[8] W. Lejkowski, A. P. Dobrowolski, K. Majka, and R. Olszewski. Spectral analysis of major heart tones [abstract]. Proc. SPIE 10715. 2017 Radioelectronic Systems Conference, 1071508 (19 April 2018).

[9] Volpicelli G, Lamorte A, Villén T. What's new in lung ultrasound during the COVID-19 pandemic. Intensive Care Med. 2020 Jul;46(7):1445-1448.

[10] Gabaldon-Figueira JC, Brew J, Doré DH, Umashankar N, Chaccour J, Orrillo V, Tsang LY, Blavia I, Fernández-Montero A, Bartolomé J, Grandjean Lapierre S, Chaccour C. Digital acoustic surveillance for early detection of respiratory disease outbreaks in Spain: a protocol for an observational study. BMJ Open. 2021 Jul 2;11(7):e051278.

[11] Hall JI, Lozano M, Estrada-Petrocelli L, Birring S, Turner R. The present and future of cough counting tools. J Thorac Dis. 2020 Sep;12(9):5207-5223.

[12] Kvapilova L, Boza V, Dubec P, et al. Continuous sound collection using Smartphones and machine learning to measure cough. Digit Biomark 2019;3:166–75.

[13] Mukundarajan H, Hol FJH, Castillo EA, et al. Using mobile phones as acoustic sensors for high-throughput mosquito surveillance. Elife 2017;6:e27854.

[14] Laguarta J, Hueto F, Subirana B. COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings. IEEE Open J Eng Med Biol. 2020 Sep 29;1:275-281.

[15] Meng L, Chan WS, Huang L, Liu L, Chen X, Zhang W, Wang F, Cheng K, Sun H, Wong KC. Mini-review: Recent advances in post-translational modification site prediction based on deep learning. Comput Struct Biotechnol J. 2022 Jun 30;20:3522-3532.

[16] Sobahi N, Atila O, Deniz E, Sengur A, Acharya UR. Explainable COVID-19 detection using fractal dimension and vision transformer with Grad-CAM on cough sounds. Biocybern Biomed Eng. 2022 Jul-Sep;42(3):1066-1080.

[17] Mantokoudis G, Kompis M, Dubach P, Caversaccio M, Senn P. How internet telephony could improve communication for hearing-impaired individuals. Otol Neurotol. 2010 Sep;31(7):1014-21.